U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 173

Data Compression Applied to NMC Grid-Point Data Fields

Harold A. Bedient
James G. Howcroft
Automation Division

MARCH 1978

DATA COMPRESSION APPLIED TO NMC GRID-POINT DATA FIELDS

Harold A. Bedient and James G. Howcroft
Automation Division, National Meteorological Center
National Weather Service, NOAA, Washington, D. C.

ABSTRACT. A data compression technique is applied to standard National Meteorological Center grid-point data fields. The delta packing method, which saves the differences between successive points instead of the point values themselves, is used.

1. INTRODUCTION

At the National Meteorological Center (NMC) the numerical weather analysis and prediction models generate values of meteorological variables at the nodes (or grid points) of a regular network over a geographical map. Every day several thousand such fields of grid-point data are generated. NMC uses these fields for further computations, to make displays and to format and transmit hundreds of grid code bulletins for world-wide distribution.

To realize economy in transmission and storage of NMC grid-point data fields, we have devised a scheme to compact them. The technique was suggested to us by Mr. B. E. Bradford of the U. S. Navy Fleet Numerical Weather Central in Monterey, California, where they have been using a similar scheme. The basic technique adapts a compaction algorithn developed by Marron and DeMaine (1967).

2. NMC STANDARD PACKING METHOD

The present standard for identifying and packing NMC grid-point data fields, described in NMC Office Note 84 (1975), prefixes each grid-point data field with an identification label that describes the contents of the field and includes scaling factors pertinent to the packed data which follows as a contiguous part of the physical record. To recover the field value (Q) in standard units, we apply multiplicative and additive constants to the 16-bit scaled integer $\hat{Q}$ as follows:

$$Q = A + (\hat{Q} * 2^{n-15})$$

where the integer $\underline{n}$ and the real constant $\underline{A}$ are found in the label.

3. SCANNING SEQUENCE

To pack the data into a linear string of numbers, we use an arbitrary sequence for scanning the two dimensional sets of grid point values. At NMC we visualize the two dimensional Q(i,j) array with the horizontal i-coordinate increasing from left-to-right and the vertical j-coordinate increasing upward. Normal scanning begins at the lower left corner, proceeds left-to-right along each row and successively upward. We will refer to this sequence as the "one direction" scanning mode.

Since the difference between the grid points at the right end of one row and the left end of the next row sometimes results in a large value, we have considered two other scanning sequences for use with the delta packing method. The "alternating direction" sequence scans one row left-to-right and the next right-to-left. The "spiral" sequence proceeds around the array spiralling inward to the last point. Figure 1 represents these three scanning sequences schematically.

Since the difference between the grid points at the right end of one row and the left end of the next row sometimes results in a large value, we have considered two other scanning sequences for use with the delta packing method. The "alternating direction" sequence scans one row left-to-right and the next right-to-left. The "spiral" sequence proceeds around the array spiralling inward to the last point. Figure 1 represents these three scanning sequences schematically.

SCANNING MODES

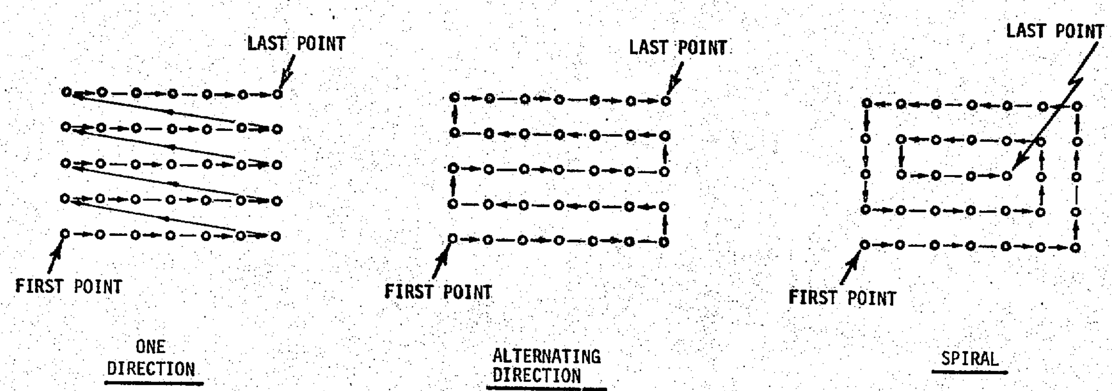

FIGURE 1

In practice we use the alternating direction scan for limited area grids and latitude-longitude grids, and the spiral scan for hemispheric polar grids. The identification label of the delta-packed data field record specifies the mode of scanning and the number of differencing passes as Table 1 shows.

Table 1. -- Mode of Scanning and Differencing for Delta Packing

Mode	Explanation	
0	One-direction	no differencing
1	One-direction	first-order differences
2	Alternating direction	first-order differences
3	Spiral	first-order differences
4	One-direction	second-order differences
5	Alternating direction	second-order differences
6	Spiral	second-order differences
7	Reserved for future use	

The identification label includes the length of the data row and the total number of grid points in the data field. Using these two values to determine the number of rows, we have the rectangular array dimensions needed for reconstructing the data field.

4. DELTA PACKING METHOD

In the delta packing method we retain the Office Note 84 label format and apply the delta packing data compression technique to the grid-point values. Since most NMC fields vary smoothly, the differences between adjacent data quantities are smaller in magnitude than the original quantities and store in fewer bits. The Navy FNWC has found that, in almost all cases, using second-order differences results in the best compaction. Our limited experience confirms this and also that, in some cases, differencing can result in expansion rather than compression. Consequently, our delta packing procedure permits the selection of either zero, one or two orders of differencing:

(a) At zero differencing passes, the array of m data points is represented by the sequence:

$$\hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \ldots, \hat{Q}_m$$

where $\hat{Q}_i$ are the scaled integer data values.

(b) With one differencing pass, the data can be represented by:

$$\hat{Q}_1, d_2, d_3, d_4, \ldots, d_m$$

where d_i is the first-order difference,

$$d_i = \hat{Q}_i - \hat{Q}_{i-1}, \text{ and } 2 \leq i \leq m$$

(c) With two differencing passes, the data can be represented by:

$$\hat{Q}_1, \delta_2, \delta_3, \delta_4, \ldots, \delta_m$$

where, δ_i is the second-order difference

$$\delta_i = d_i - d_{i-1}, \text{ and } 2 \leq i < m$$

Note that the first data point has its scaled integer representation preserved and that, after a second differencing pass,

$$\delta_2 = d_2 - \hat{Q}_1 = \hat{Q}_2 - 2\hat{Q}_1$$

After we form the differences, we scan each data element to determine the number of bits needed to store the data. We consider partial sequences to find the number of differences (D) that fit into a string of 64 bits with the 8-leftmost (high order) bits reserved for the value of D and the differences packed into the remaining 56 bits. Each difference occupies a bit string whose length is the integer quotient found by dividing 56 by D. The remainder indicates the number of unused bits located in the right (low order) portion of the 64 bit string. This packing procedure continues until it exhausts the entire sequence of differences. If the final packing requires only half of a 64 bit string, we omit the low order 32 bits.

Table 2 summarizes the various combinations of the number of data differences, the bit-string lengths, and the number of unused low-order bits that occur in the delta packing process. A special case occurs when the packer encounters a string of more than 28 zeroes to be packed into the 64-bit string. In this case, the negative value of D is stored in a 32 bit string. Only those 32 bits are kept; the low order half of the 64 bit string is omitted.

Table 2. Structure of Packing in a 64-Bit String

D	Bit String Length	Unused low order bits
3, 2, 1	16	8, 24, 8
4	14	0
5	11	1
6	9	2
7	8	0
8	7	0
9	6	2
10, 11	5	6, 1
12, 13, 14	4	8, 4, 0
15, 16, 17, 18	3	11, 8, 5, 2
19 thru 28	2	(56 - 2*D)
>28	Special case, see text	

5. RECOVERING FROM DELTA PACKING

The data record label contains the information needed to restructure a data field from its delta packed values. Attachment 1, a modification to Appendix C of Office Note 84, describes the label for the delta packed data fields. The first eight words (32 bit strings) of the label are unchanged. Word 9 contains adjusted values for the number of bytes in the record and the checksum, which is a "exclusive or" of all the other 16 bit halfwords in the record.

Additionally, the scaling value n in the low order half of word 11 is adjusted for any rescaling performed to remove unnecessary data significance. This rescaling will be transparent to the user.

Word 12 contains the scan mode, described in Table 1, in bits 0-3 and the number of points on a horizontal i-row in the bits 4-15. The low order bits of word 12 contain the scaled integer value of the last point of the packed array for comparison with the last point unpacked by the user. They should agree.

To illustrate the technique of restoring from delta packing, suppose that the first several words of a delta field are (in hexadecimal):

0080810C	00000000	00000000	00000000
000000FF	00000000	4E010300	002902C0
0310BBCF	43377A80	0000000D	502001E5
05419800	037DFFFE	072000E0	020228B0
05DF1E24	E904DEBC	etc.	

The first 12 words are the data field identification. O. N. 84 reveals that this is a forecast of an earth's surface pressure field valid 12 hours after 00Z on January 3, 1978. The grid type is shown as FF, but by prior agreement with the user is known to be a selected rectangular subset of the LFM grid. There are 704 data points ($2C0_{hex}$) and the length of a horizontal row is 32 grid points (20_{hex}); consequently, the subset has dimensions 32 by 22. The scanning mode (5) indicates an alternating direction row scan with two differencing passes.

The last point in the array -- which one can determine has coordinates (i,j) = (1,22) -- has the scaled integer representation $\hat{Q} = 1E5_{hex}$, equivalent to decimal 485. Using the mid-range value of $\underline{A}$ = 887.656 ($43377A80_{hex}$) and scale factor $\underline{n}$ = 13 (D_{hex}), a physical value of 1008.9 mb for that particular point is determined.

Words 13 and 14 provide the first 64 bit sequence of the delta packed data points. There are 5 grid points involved, each contained in 11 bits with a zero fill in the right-most bit. After sorting out the bits, we are presented with the values (in hex) of

20C 600 006 7DF 7FF

which convert to signed integers (hex)

20C	-200	6	-21	-1

and to decimal integers

$\hat{Q}_1$	δ_2	δ_3	δ_4	δ_5
524	-512	6	-33	-1

After restoring one delta pass, the following values result

$\hat{Q}_1$	d_1	d_2	d_3	d_4
524	12	18	-15	-16

and after another

$\hat{Q}_1$	$\hat{Q}_2$	$\hat{Q}_3$	$\hat{Q}_4$	$\hat{Q}_5$
524	536	554	539	523

These convert to physical values, to the nearest tenth of a millibar,

Q_1	Q_2	Q_3	Q_4	Q_5
1018.7	1021.7	1026.2	1022.4	1018.4

for the first five grid points of the array.

REFERENCES

Marron, B. A., and P. A. D. deMaine, 1967: Automatic Data Compression, Communications of the Association for Computing Machinery, vol. 10, pgs. 711 - 715, (November 1967).

NMC Office Note 84, 1975: Labels for NMC 360/195 Data Fields, revised June 1975.

ATTACHMENT 1

(A Modification of Appendix C of NMC Office Note 84)

0 4 8 12 16 24 31

Word				
1	Q Data Type (12)	S1 Type of Surface 1 (12)	F1 Time 1 (8)	
2	T Mark (4)	C1 Numerical Value of Surface 1 (20)	E1 Exponent 1 (8)	
3	M Mark (4)	X Exception Marker (8)	S2 Type of Surface 2 (12)	F2 Time 2 (8)
4	N Mark (4)	C2 Numerical Value of Surface 2 (20)	E2 Exponent 2 (8)	
5	CD (8)	CM (8)	KS Gridding Method (8)	K Grid Type (8)
6	Unused (4)	RN Record Number on Disk (12)	NW Number of Words in Record (16)	
7	YY Year (8)	MM Month (8)	DD Day (8)	II Initial Hour (8)
8	R Run Marker (8)	G Generating Program (8)	J Number of Data Points (16)	
9	B Number of Bytes in Record (Adjusted Delta Count) (16)	Z Checksum (Adjusted for Delta Packing) (16)		
10	A Mid-range Value, Real*4 (32)			
11	Unused (16)	n Scaling Value, Integer*2 (Adjusted for Delta Packing) (16)		
12	Delta Mode (4)	LI Length of Row (12)	Last Data Word in Original Form (16)	
13	Data Point Values ...			